Übungsaufgaben
Automaten und Formale Sprachen

Universität München, CIS, WS 2003/2004

Abgabe: Do 15.02.03

Aufgabe 6.1 (Pumping-Lemma für lineare Sprachen) Zu jeder linearen Sprache $L \subseteq \Sigma^*$ gibt es eine Zahl n_L, so daß jedes Wort $w \in L$ mit $|w| > n_L$ eine Zerlegung der Form $w = xuyvz$ hat mit:

(a) $uv \neq \epsilon$,

(b) $|xz| \leq n_L$,

(c) $\{ xu^nyv^nz \mid n \in \mathbb{N} \} \subseteq L$.

Sei $\Sigma := \{a, b\}$ mit $a \neq b$. Die Sprache

$$L := \{ a^nb^n \mid n \in \mathbb{N} \} \quad \subseteq \Sigma^*$$

ist linear: sie ist (in 2^{Σ^*}) die kleinste Lösung x der Gleichung $x = axb + 1$.

(a) Zeige mit dem Pumping-Lemma für lineare Sprachen, daß die Sprache $L^2 := LL$ nicht linear ist. (Das Sprachprodukt linearer Sprachen ist also i.a. keine lineare Sprache.) (4 Punkte)

(b)* Eine Sprache heißt *m-linear*, wenn sie das Produkt $L_1 \cdots L_m$ von m linearen Sprachen $L_1, \ldots, L_m$ ist. Da $\{\epsilon\}$ eine lineare Sprache ist, ist jede m-lineare Sprache auch $(m + 1)$-linear.
Können Sie zeigen, daß die 3-lineare Sprache $L^3 := LLL$ nicht 2-linear ist? (4 Punkte)

Aufgabe 6.2 (Unzusammenhängende Konstituenten) In natürlichen Sprachen gibt es Ausdrücke, die in den Sätzen der Sprache nicht immer als zusammenhängende Teilfolge der Wortfolge des Satzes auftreten, z.B. im Deutschen bei

(a) Nominalphrasen mit Relativsätzen: *Sie hat <u>das Buch, das ich ihr schenkte</u>, gelesen.* und *Sie hat <u>das Buch</u> gelesen, <u>das ich ihr schenkte</u>.*

(b) Adjektivphrasen mit Attributen: *Das Gen ist <u>nützlich für die Fortplanzung</u>.* und *ein <u>nützliches</u> Gen <u>für die Fortpflanzung</u>.*

Man kann daher die Nominal- bzw. Adjektiv-Phrasen als i.a. *gespaltene Ausdrücke* (*split strings*) auffassen, in die andere Ausdrücke *eingefügt* werden können. Der *Einfügung* des Worts $w \in \Sigma^*$ in das gespaltene Wort $(u, v) \in \Sigma^* \times \Sigma^*$ sei mit $(u, v) : w \;:= uwv \in \Sigma^*$ bezeichnet. Insbesondere ist $uv = (u, v) : \epsilon$. Das kann man auf Mengen von gespaltenen und einfachen Wörtern ausdehnen durch

$$R : L := \{ (u, v) : w \mid (u, v) \in R, w \in L \}.$$

Die Kategorien einer Grammatik bezeichnen Wortmengen (z.B. bei den Wortarten N, A, V usw.) oder Mengen gespaltener Ausdrücke (z.B. bei NP, AP). Die Grammatik definiert diese Kategorien rekursiv durch einander, z.B. (mit abnehmender Bindungsstärke der Operatoren $:, \cdot, \times, \mid$ (bzw. $+$))

$$\begin{array}{rcl}
S^{zweit} &=& NP : 1 \cdot Aux \cdot NP : V^{part2} \mid \cdots \\
NP &=& Pron \times 1 \mid Det \cdot AP : N \times (Rel \mid 1) \\
AP &=& A \times (PP \mid 1) \\
Rel &=& RelPron \cdot NP : 1 \cdot V^{part2} \cdot Aux \mid \cdots \\
PP &=& Präp \cdot NP : 1 \mid \cdots
\end{array}
\qquad
\begin{array}{rcl}
Det &=& ein \mid das \mid \cdots \\
N &=& Gen \mid \cdots \\
Pron &=& sie \mid \cdots \\
A &=& nützlich \mid \cdots \\
V^{part2} &=& geschenkt \mid \cdots
\end{array}$$

(a) Kann man mit diesen Regeln (und geeignetem Lexikon) einen Satz wie *Der Wissenschaftler hat ein nützliches Gen für die Fortpflanzung vermarktet, das er endeckt hat.* aufbauen, und wie sähen die Aufbauschritte aus ? (3 Punkte)

(b) Stelle den Aufbau durch einen Baum dar, dessen Blätter wie die Wörter angeordet sind, und dessen Knoten mit den Kategorien der jeweiligen (nicht notwendig zusammenhängenden) Teilausdrücke markiert sind. (3 Punkte)